Peramalan Kunjungan Wisata dengan Pendekatan Model SARIMA (Studi kasus : Kusuma Agrowisata)

Nofinda Lestari dan Nuri Wahyuningsih
Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111
E-mail: nuri@matematika.its.ac.id

***Abstrak*—Peramalan jumlah kunjungan wisata yang masuk ke dalam suatu daerah sangat diperlukan oleh pelaku bisnis pariwisata. Untuk itu tujuan utama dalam penelitian ini adalah untuk pembentukan model dan memperoleh hasil peramalan jumlah kunjungan wisata satu periode ke depan dengan studi kasus di Kusuma Agrowisata Batu Malang. Metode yang digunakan dalam penelitian ini adalah metode Box Jenkins dengan pendekatan model SARIMA sebagai pengembangan dari model ARIMA. Metode ini sesuai dengan situasi dengan data yang ada bersifat musiman. Langkah pertama yang dilakukan adalah melihat kestasioneran data. Selanjutnya identifikasi model dari perhitungan ACF dan PACF. Dari perhitungan ACF dan PACF bisa dibentuk model ARIMA sementara, kemudian penaksiran dan estimasi parameter model, dan langkah yang terakhir adalah pemeriksaan diagnostik dengan melihat hasil residual dan normalitas. Hasil analisis menunjukkan bahwa model ARIMA ([2,5],1,1)(1,0,0)^{12} adalah model yang terbaik.**

***Kata Kunci*—ARIMA, Kunjungan Wisata, Metode Box-Jenkins, Model SARIMA**

I. PENDAHULUAN

PARIWISATA merupakan bagian yang tidak terpisahkan dari kehidupan manusia. Indonesia sebagai negara yang sedang berkembang dalam tahap pembangunannya, berusaha membangun industri pariwisata sebagai salah satu cara untuk mencapai neraca perdagangan luar negeri yang berimbang [1].

Setiap kota di Indonesia memiliki objek wisatanya masing-masing, Kota Malang, Jawa Timur sebagai salah satu tujuan wisata di Jawa Timur semakin populer. Selain karena udaranya yang sejuk, banyak juga tujuan wisata di sekitar Malang, seperti Kota Batu yang mempunyai beragam wisata alam, contohnya Kusuma Agrowisata. Oleh karena itu penelitian ini bertujuan untuk meramalkan jumlah pengunjung Kusuma Agrowisata pada tahun 2012 [2].

Pada musim–musim liburan sekolah dan juga tahun baru, jumlah kunjungan wisata di berbagai tempat pariwisata lebih meningkat daripada hari-hari biasa, sehingga model peramalan yang digunakan dalam penelitian ini adalah model SARIMA. Beberapa penelitian terdahulu memberikan informasi mengenai metode yang dapat diterapkan untuk meramalkan jumlah kunjungan wisata. *Naïve Models*, *Exponential Smoothing* dan *Seasonal ARIMA* merupakan metode yang pernah digunakan oleh Wang dan Lim untuk meramalkan kunjungan wisatawan ke Malaysia dan asal Australia ke Jepang. Sedangkan Athanasopoulos dan Hyndman memanfaatkan *Time Series Regression, Exponential Smoothing* dan *Innovations state space models* dalam meramalkan jumlah kunjungan wisata ke Australia.

Model SARIMA juga pernah digunakan oleh Hermanto untuk meramalkan tingkat penjualan motor di PT. Lancar Sukses Mandiri, oleh Adit untuk memprediksi jumlah wisatawan dalam negeri yang datang ke hotel di Daerah Istimewa Yogyakarta, oleh Roni untuk meramalkan harga bawang merah di enam kota besar Indonesia, oleh Anggraini untuk memodelkan SARIMA dan penerapannya [3, 4, 5, 6].

II. URAIAN PENELITIAN

A. *Studi Literatur*

Metode Peramalan adalah cara memperkirakan secara kuantitatif apa yang akan terjadi pada masa yang akan datang, berdasarkan data yang relevan pada masa lalu. Metode ini sangat berguna dalam mengadakan pendekatan analisis terhadap perilaku atau pola dari data yang lalu, sehingga dapat memberikan cara pemikiran, pengerjaan dan pemecahan yang sistematis dan prakmatis serta memberikan tingkat keyakinan yang lebih. Salah satu metode dalam peramalan yaitu metode Box Jenkins. Beberapa model dalam Metode Box-Jenkins yaitu :

1) Model ARIMA (p,d,q)

Rumus umum model ARIMA (p,d,q) adalah sebagai berikut [7]:

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B) a_t$$

dengan :

$\phi_p(B) = (1-\phi_1 B - \ldots - \phi_p B^p)$, AR (p)

$\theta_q(B) = (1-\theta_1 B - \ldots - \theta_q B^q)$, MA (q)

$(1-B)^d$: *differencing* orde d

a_t : nilai residual pada saat t

2) Model ARIMA dan Faktor Musim (SARIMA)

Notasi ARIMA dapat diperluas untuk menangani aspek musiman, notasi umumnya adalah [7] :

ARIMA (p,d,q) (P,D,Q)^S

dengan :

p,d,q : bagian yang tidak musiman dari model

$(P,D,Q)^S$: bagian musiman dari model
S : jumlah periode per musim

Adapun rumus umum dari ARIMA (p,d,q)$(P,D,Q)^S$ sebagai berikut :

$$\Phi_P(B^S)\phi_p(B)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\Theta_q(B^S)a_t$$

dengan :

$\phi_p(B)$: *AR non seasonal*

$\Phi_P(B^S)$: *AR seasonal*

$(1-B)^d$: *differencing non seasonal*

$(1-B^S)^D$: *differencing seasonal*

$\theta_q(B)$: *MA non seasonal*

$\Theta_Q(B^S)$: *MA seasonal*

3) Stasioneritas data

Kestasioneran data bisa dilihat dari plot time series. Untuk melihat kestasioneran data dalam means bisa dilihat dari perhitungan ACF dan PACF nya. ACF untuk data jumlah kunjungan wisata diperoleh dengan rumus sebagai berikut :

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{k=1}^{n-k}\left(Z_t-\bar{Z}\right)\left(Z_{t+k}-\bar{Z}\right)}{\sum_{t=1}^{n}\left(Z_t-\bar{Z}\right)^2} \qquad (1)$$

dengan Z_t data time series pada waktu ke t dan $\bar{Z}$ rata-rata sampel. Sedangkan PACF untuk data jumlah kunjungan wisata diperoleh dengan rumus sebagai berikut :

$$\hat{\phi}_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1}\hat{\phi}_{k-1,j}\,\rho_{k-j}}{1-\sum_{j=1}^{k-1}\hat{\phi}_{k-1,j}\,\rho_j} \qquad (2)$$

dengan ρ_k adalah fungsi autokorelasi.

Ketidakstasioneran data dalam means dapat diatasi dengan proses pembedaan *(differencing)*, sedangkan kestasioneran data dalam varians dapat dilihat dengan nilai λ. Adapun nilai λ dihitung dengan rumus sebagai berikut :

$W_i = \frac{Y_i^{\lambda}-1}{\lambda G^{\lambda-1}}$ untuk $\lambda \neq 0$, $W_i = G\ln\left(Y_i\right)$ untuk $\lambda = 0$, sehingga

$$MR_i = \max\left[W_i,\ldots,W_{i-r+1}\right]-\left[W_i,\ldots,W_{i-r+1}\right]$$

$$\overline{MR} = \frac{(MR_i+\ldots+MR_N)}{(N-r+1)} \qquad (3)$$

dengan, Y_i data aktual untuk $i=1,\ldots,n$, G *geometric mean* dari seluruh data, λ nilai lambda, n jumlah data observasi.

4) Penaksiran dan Pengujian Parameter

Model peramalan yang diperoleh akan diuji signifikansi parameter modelnya dengan hipotesis sebagai berikut.

Hipotesa :

H_0 : estimasi parameter = 0

H_1 : estimasi parameter $\neq 0$

Statistik uji : $t_{hitung} = \dfrac{\text{estimasi parameter}(\phi_1)-0}{\text{standart error parameter}}$

Daerah penolakan : Tolak H_0 jika $|t_{hitung}| > t_{\alpha/2-n-1}$

dimana n menunjukkan banyaknya data dan n menunjukkan banyaknya parameter.

Kriteria pengujian:

Jika $|t_{hitung}| > t_{\alpha/2-n-1}$ maka H_0 ditolak.

5) Uji Asumsi Residual

Dalam menentukan model ARIMA yang terbaik, harus dipilih model yang seluruh parameternya signifikan, kemudian juga memenuhi 2 asumsi residual yaitu berdistribusi normal dan *white noise.*

a. Distribusi Normal

Pengujian kenormalan dapat dihitung dengan menggunakan Kolmogorov-Smirnov.

Hipotesa :

H_0 : residual berdistibusi normal

H_1 : residual tidak berdistribusi normal

Statistik uji : $D_{hit} = \sup_x |S(x)-F_0(x)|$

Kriteria pengujian:

Jika $D_{hit} > D_{(1-\alpha),n}$ maka H_0 ditolak.

dengan :

$F_0(x)$:fungsi yang dihipotesiskan yaitu berdistribusi normal

$S(x)$:fungsi distribusi kumulatif dari data asal

n :banyaknya residual

$D_{(1-\alpha),n}$ didapatkan dari Tabel Kolmogorov Smirnov.

b. White Noise

Suatu model bersifat *white noise* artinya residual dari model tersebut telah memenuhi asumsi identik (variasi residual homogen) serta independen (antar residual tidak berkorelasi). Pengujian asumsi *white noise* dilakukan dengan menggunakan uji Ljung-Box.

Hipotesa :

H_0 : $\rho_1 = \rho_2 = \ldots = \rho_k = 0$

H_1 : Minimal ada satu ρ_i yang tidak sama dengan nol, $i = 1.2....,k$

Statistik uji : $Q = n(n+2)\sum_{k=1}^{K}\frac{\hat{\rho}_k^2}{n-k}, n > k$

Daerah penolakan : $Q > \chi^2_{(\alpha;K-p-q)}$

dengan :

K : lag maksimum
n : jumlah data (observasi)
k : lag ke-k
p dan q : order dari ARMA (p,q)
$\hat{\rho}_k$: autokorelasi residual untuk lag ke-k

Salah satu prosedur pemeriksaan diagnosis yang dikemukakan Box Jenkins adalah overfitting, yakni dengan menambah satu atau lebih parameter dalam model yang dihasilkan pada tahap identifikasi. Karena ada salah satu estimasi parameter yang tidak signifikan maka dilakukan tahap overfitting. Model yang dihasilkan dari hasil overfitting dijadikan sebagai model alternatif yang kemudian dicari model yang terbaik diantara model-model yang signifikan.

6) Kriteria Pemilihan Model Terbaik

Dalam pemilihan model terbaik kadang mudah namun dilain waktu pemilihan modelnya kadang menjadi lebih sulit. Oleh karena itu dibutuhkan kriteria untuk menentukan model yang terbaik dan akurat. pemilihan model terbaik dapat menggunakan *Mean Absolute Percentage Error* (MAPE) yaitu :

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{n} \times 100\%$$

n menyatakan banyaknya data yang akan dihitung residualnya. Semakin kecil nilai RMSE dan MAPE, maka semakin baik dan model tersebut layak untuk digunakan. Adapun nilai MSE dan RMSE diperoleh dengan rumus sebagai berikut :

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_t - \hat{y}_t \right)}$$

B. *Sumber Data*

Data yang dipakai dalam penelitian ini adalah data kunjungan wisata Kusuma Agrowisata Batu Malang mulai tahun 2001-2011 yang didapat dari laporan pihak marketing Kusuma Agrowisata Batu Malang.

C. *Metodologi Penelitian*

Identifikasi data dilakukan dengan membuat plot time series untuk melihat apakah ada indikasi data musiman atau tidak dan melihat data sudah stasioner atau belum. Jika data belum stasioner dalam *mean* dilakukan *differencing*, namun jika data tidak stasioner dalam varians, dilakukan transformasi Box-Cox. Kemudian menghitung ACF dan PACF dari data yang sudah stasioner. Dari perhitungan ACF dan PACF yang didapat bisa ditentukan model ARIMA sementara.

Tahap selanjutnya adalah penaksiran dan pengujian parameter ¯$_q$ dan ¤$_p$. Pada pengujian parameter ¯$_q$,¤$_p$ dicek apakah parameter yang didapat dari model SARIMA sementara signifikan atau tidak.

Tahap selanjutnya adalah pemeriksaan diagnostik. Pemeriksaan diagnostik dilakukan untuk uji identik, uji independent, uji normalitas, dan *overfitting*. Nilai residual dari model kemudian diuji *whitenoise* dan kenormalan distribusinya. Jika pada tahap ini didapatkan kesimpulan bahwa model belum sesuai maka prosedur ini kembali ke tahap penaksiran dan pengujian parameter.

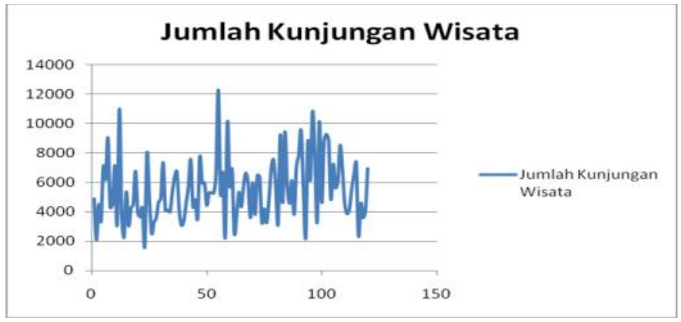
Langkah terakhir adalah pemilihan model terbaik. Model terbaik dipilih dari model-model yang telah memenuhi semua asumsi yaitu parameter signifikan, *white noise*, dan berdistribusi normal, serta nilai MAPE terkecil.

III. HASIL PENELITIAN

Dasar dari pendekatan metode Box-Jenkins dibagi menjadi 3 tahap.

A. *Tahap Identifikasi*

Yang pertama kali dilakukan yaitu membuat plot data time series seperti ditunjukkan pada Gambar 1. Namun setelah dilihat pada Gambar 1 tidak bisa dipastikan apakah data jumlah kunjungan wisata Kusuma Agrowisata termasuk musiman atau tidak.



Gambar. 1. Plot time series data jumlah kunjungan wisata.

Tabel 1.
Perhitungan ACF dan PACF

Lag	°$_k$	¤̂$_{kk}$
1	-0.554006*	-0.554006*
2	0.166143	-0.203122*
3	-0.135652	-0.206841*
4	-0.139535	-0.461579*
5	0.195992	-0.301312*
6	-0.041697	-0.201342*
7	0.202131	0.135513
8	-0.236899	0.003565
9	0.051972	-0.003714
10	-0.042439	0.069203
11	-0.216270	-0.458017*
12	0.481885*	-0.115634
13	-0.206168	0.140560
14	0.071262	0.090881
15	-0.145338	-0.020356
16	-0.0412832	0.056475
17	0.0114817	-0.126167
18	0.110073	-0.142088
19	0.145473	0.034267
20	-0.234585	-0.036337
21	0.105491	-0.016161
22	-0.243787	-0.332464*
23	0.181348	0.007778
24	0.0868786	0.107872
25	-0.001549	-0.055735
26	-0.020498	-0.111093
27	-0.098928	0.078717
28	0.027729	0.070452
29	-0.102281	-0.084433
30	0.187224	-0.057767

Keterangan : * adalah lag yang keluar dari batas normal.

Setelah membuat plot time series langkah selanjutnya adalah melihat kestasioneran data dalam varians dan mean.

Dengan menggunakan rumus pada persamaan (3) untuk data kunjungan pariwisata kusuma agrowisata diperoleh nilai $\lambda = 1$. Oleh karena itu data kunjungan wisata sudah stasioner dalam varians. Setelah data sudah stasioner dalam varians maka langkah selanjutnya adalah melihat kestasioneran data dalam means. Dengan menggunakan persamaan (1) dan (2) diperoleh hasil perhitungan ACF dan PACF. Adapun hasil perhitungan ACF dan PACF dari data yang sudah stasioner dalam varians seperti dilihat pada Tabel 1. Dari Tabel 1 model ARIMA sementara untuk data kunjungan wisata adalah ARIMA.

B. Tahap Penaksiran dan Pengujian

Pengujian signifikansi parameter model dengan $\alpha = 5\%$ dan menggunakan uji-t adalah sebagai berikut :

1) Pengujian parameter ϕ_1

Model peramalan yang diperoleh dari ARIMA sementara diuji signifikansi parameter ϕ_1 dengan hipotesis sebagai berikut.

Hipotesa :

H_0 : estimasi parameter $\phi_1 = 0$

H_1 : estimasi parameter $\phi_1 \neq 0$

Statistik uji :

$$t_{hitung} = \frac{\text{estimasi parameter}(\phi_1) - 0}{\text{standart error parameter}} = \frac{0.80820}{0.03267} = 24.74$$

$$t_{tabel} = t_{\frac{\alpha}{2}, n-1} = t_{0.025:119} = 1.960$$

Karena $|t_{hitung}| > t_{\frac{\alpha}{2}, n-1}$ maka H_0 ditolak. Sehingga dapat dikatakan estimasi parameter ϕ_1 signifikan.

2) Pengujian parameter ϕ_2

Model peramalan yang diperoleh dari ARIMA sementara diuji signifikansi parameter ϕ_2 dengan hipotesis sebagai berikut.

Hipotesa :

H_0 : estimasi parameter $\phi_2 = 0$

H_1 : estimasi parameter $\phi_2 \neq 0$

Statistik uji :

$$t_{hitung} = \frac{\text{estimasi parameter}(\phi_2) - 0}{\text{standart error parameter}} = \frac{-0.25666}{0.03733} = -6.88$$

$$t_{tabel} = t_{\frac{\alpha}{2}, n-1} = t_{0.025:119} = 1.960$$

Karena $|t_{hitung}| > t_{\frac{\alpha}{2}, n-1}$ maka H_0 ditolak. Sehingga dapat dikatakan estimasi parameter ϕ_2 signifikan.

2) Pengujian parameter Θ_1

Model peramalan yang diperoleh dari ARIMA sementara diuji signifikansi parameter Θ_1 dengan hipotesis sebagai berikut.

Hipotesa :

H_0 : estimasi parameter $\Theta_1 = 0$

H_1 : estimasi parameter $\Theta_1 \neq 0$

Statistik uji :

$$|t_{hitung}| = \frac{\text{estimasi parameter}(\Theta_1) - 0}{\text{standart error parameter}} = \frac{0.24326}{0.09333} = 2.61$$

$$t_{tabel} = t_{\frac{\alpha}{2}, n-1} = t_{0.025:119} = 1.960$$

Karena $|t_{hitung}| > t_{\frac{\alpha}{2}, n-1}$ maka H_0 ditolak. Sehingga dapat dikatakan estimasi parameter Θ_1 signifikan. Adapun hasil pengujian parameter Θ_1 didapat hasil seperti pada Tabel 2.

Tabel 2.
Pengujian Parameter

Model	Parameter	Estimasi	Std
(5,1,[1,12])	ϕ_1	0.80820	0.03267
	ϕ_2	-0.25666	0.03733
	Θ_1	0.24326	0.09333

C. Pemeriksaan Diagnostik

Model dikatakan sudah sesuai apabila telah memenuhi uji-uji sebagai berikut :

1) Uji identik

Karena pada tahap identifikasi Y_t sudah stasioner dalam means dan varians, maka model dapat dikatakan sudah identik.

2) Uji independent

Pengujian ini dilakukan untuk melihat adanya independensi dari residual (untuk mengetahui korelasi dari nilai-nilai residual) maka dilakukan uji :

Hipotesa:

$H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0$

H_1 : Minimal ada satu ρ_i yang tidak sama dengan nol, $i = 1.2....,k$

Statistik uji : $Q = n(n+2)\sum_{k=1}^{K} \frac{\hat{\rho}_k^2}{n-k}, n > k$

Untuk $K = 6$ maka :

$$Q = 120(120+2)\sum_{k=1}^{6} \frac{\hat{\rho}_k^2}{120-k}$$

$$Q = 14640\left(\frac{(-0.150)^2}{120-1}+\frac{(-0.026)^2}{120-2}+\frac{(-0.143)^2}{120-3}+\frac{(-0.091)^2}{120-4}+\frac{(0.031)^2}{120-5}+\frac{(0.234)^2}{120-6}\right)$$

$$= 13.46$$

$$\text{i}^{2}_{(\sim;K-p-q)} = \text{i}^{2}_{(0.05;3)} = 7.81473$$

Karena $Q > \text{i}^{2}_{(\sim;K-p-q)}$ maka H_0 ditolak artinya residual tidak *white noise*. Dengan cara yang sama bisa dipakai untuk melihat adanya independensi dari residual pada lag 12, 18, dan 24.

3) Uji normalitas

Untuk melihat apakah residual berdistribusi normal. Maka dilakukan uji:

Hipotesa:

H_0 : residual berdistibusi normal

H_1 : residual tidak berdistribusi normal

Statistik uji :

$$D_{hit} = \sup_x |S(x) - F_0(x)|$$

$$= 0.048717$$

$$D_{(1-\sim,n)} = 0.111$$

Karena $D_{hit} \le D_{(1-\sim,n)}$ maka H_0 diterima artinya residual berdistribusi normal.

Beberapa model alternatif yang kemudian dicari model yang terbaik diantara model-model yang signifikan.

Adapun model-model alternatif yang akan diujikan seperti pada Tabel 4.

Tabel 3.
Pengujian Model SARIMA

Model	Keputusan		
	Uji Parameter	Uji White Noise	Uji Normal
$([1,5],1,[1,12])(1,0,0)^6$	Signifikan	Tidak White Noise	Normal
$([2,5],1,1)(1,0,0)^6$	Signifikan	Tidak White Noise	Normal
$(1,1,[2,5])(1,0,0)^{12}$	Signifikan	White Noise	Tidak Normal
$([2,5],1,1)(1,0,0)^{12}$	Signifikan	White Noise	Normal
$([2,5],1,1)(0,0,1)^{12}$	Signifikan	Tidak White Noise	Normal
$(5,1,1)(0,0,1)^{12}$	Signifikan	Tidak White Noise	Normal

Dari Tabel 4 didapatkan satu model yang memenuhi semua asumsi yaitu model ARIMA $([2,5],1,1)(1,0,0)^{12}$. Dengan hasil perhitungan MSE, RMSE dan MAPE pada Tabel 4.

Tabel 4.
Perhitungan MSE, RMSE dan MAPE

Residual		MAPE(%)
MSE	RMSE	
882175.4	939.2419	15.93689

IV. KESIMPULAN

Model yang sesuai untuk kunjungan wisata Kusuma Agrowisata Batu Malang adalah

$$(1-0.28436B^2 - 0.31390B^5)(1-0.58881B)(1-B)Y_t = (1-B)a_t$$

dengan :

$$Y_t = [\ln(Z_t)]^{2.99}$$

DAFTAR PUSTAKA

[1] Anonim 1. Bab 2 Landasan Teori, di akses pukul 22.24 WIB, tanggal 19 Februari 2012. <http://repository.usu.ac.id/bitstream/123456789/23541/4/ Chapter%20II.pdf >..

[2] Anonim 2. Tips Wisata ke Kota Malang, , di akses pukul 09.50 WIB tanggal 28 Februari 2012. <http://travel.kompas.com/read/2011/11/13/ 17384179/Tips.Wisata.ke.Kota.Malang>.

[3] Hermanto. "Analisis Peramalan Tingkat Penjualan Motor Menggunakan Metode SARIMA Berdasarkan Pola Data Seasonality Pada PT. Lancar Sukses Mandiri". Tugas Akhir, Jurusan Teknik Informatika dan Statistika, Bina Nusantara. (2007).

[4] Adit, " Prediksi Jumlah Wisatawan Dalam Negeri Yang Datang Ke Hotel Di Daerah Istimewa Yogyakarta". Tugas Akhir, Jurusan Statistika, (2007).

[5] Roni, I.K. "Peramalan Dan Faktor-Faktor Yang Mempengaruhi Harga Bawang Merah Enam Kota Besar Indonesia". Tugas Akhir, Jurusan Ekstensi Manajemen Agribisnis, IPB. (2007).

[6] Anggaini. "Model SARIMA Dan Penerapannya". Tugas Akhir, Jurusan Matematika, UNY. (2009).

[7] Wei, W.W.S. "Time Series Analysis, Univariate and Multivariate Methods". 2nd edition. Pennsylvania: Pearson Education Inc. 2006.